# First experiences in implementing a spatial metadata repository

**Sophie Cockcroft**
Department of Information Science
University of Otago
PO Box 56
Dunedin
E-mail: Scockcroft@commerce.otago.ac.nz

## Abstract

Integrated software engineering environments (ISEE) for traditional non spatial information systems are well developed, incorporating Database Management Systems (DBMS) and Computer Aided Software Engineering (CASE) tools. The core component of the ISEE is the repository. It brings all the other components together and provides a common area to which all tools can link. In this fashion it also provides a central point for control. No such facility exists for the management of spatial data.
This paper describes the development of such a facility in the form of a spatial metadata repository

## 1. Introduction

Whilst many Spatial Information Systems (SIS) incorporate a data dictionary, it is typically a passive catalogue of features, attributes and descriptions. For example ESRI (Environmental Systems Research Institute 1995) defines the data dictionary as follows:

*"A data dictionary is a list that maintains for each coverage the names of the attributes, and a description of the attribute values (including a description of each code if necessary). Having a data dictionary for your database is invaluable as a reference during the project as well as for transferring information to others"*

This definition does not suggest an active data dictionary with the potential to impose integrity constraints. In contrast to data dictionaries in non spatial information systems the emphasis, in this definition, is on the dictionary as a tool for reference.

The term repository describes a database facility which has evolved, in mainstream database technology, from the type of passive data dictionary described above. The role of the repository is twofold; firstly the cataloguing of physical database objects through data about data or metadata during system development and secondly actively managing database integrity control during production. The means by which integrity control is managed is described in section 3.

The concept of metadata has evolved through several disciplines. At its simplest level metadata is additional information necessary for data to be useful. A more insightful explanation was provided by Henderson (Henderson 1987) who classified metadata into *dictionary* metadata describing characteristics, relationships and uses and *directory* metadata describing where the data is and how it can be accessed. Both types of metadata have received much attention from the Geographical Information Systems (GIS) community recently (Medyckyj-scott et al. 1996)(FGDC 1994). The emphasis, however tends to be on directory metadata. It is dictionary metadata which is of interest here.

## 2. Problem definition

The inability of current spatial information systems to enforce integrity constraints poses a serious threat to the quality of data entered into such systems. This inability, and suggestions for addressing it, are recurring themes in the database and GIS literature.

Worboys et al (Worboys 1994) acknowledge that the majority of GIS implementing the relational model use a hybrid architecture where attribute data is maintained in a conventional DBMS, but spatial data is organised and manipulated using conventional file handling techniques. The major reason cited for the inability to provide basic DBMS functionality, such as integrity control, within Spatial Information Systems (SIS) was the fact that the entire map base is not maintained within a single DBMS. The emphasis placed on the shortcomings of the hybrid approach suggests that one of the major integrity control problems encountered was maintaining the currency of linkages between two databases. For example GIS of this type will typically have the problem that updates on spatial databases result in updates on attribute data but not vice versa. This is particularly important in an enterprise wide situation where other information systems may request updates on the GIS. This is a problem which is rapidly being addressed in commercial DBMS. A further discussion of this issue appears in section 4.

Worboys et al (Worboys 1994) also provided a review of object oriented solutions to the problem of disparate spatial and non spatial databases within SIS. It becomes apparent from this review that there are degrees of integrity control ranging from maintaining the currency of linkages between spatial and non-spatial data, which may result from the inadequacies of spatial databases described above, to more subtle user defined integrity requirements. For example it is not usually possible for a user to specify that a value must be within a certain range or that it is only valid in connection with other values. Gunther et al (Günther and Lamberts 1994) focussed on the latter problem noting that Geographic Information Systems lack the functionality to preserve semantic integrity. By contrast in non-spatial DBMS it is commonly possible to maintain consistency according to user defined constraints. The solution proposed by Gunther et al (Günther and Lamberts 1994,Günther, 1993 #231) was also an object oriented one. Integrity constraints were, however, cited as an optional feature to be added once basic DBMS functionality was established.

Medeiros et al (Medeiros and Pires 1994) cited the problem of enforcing spatial integrity constraints as one which derives from the existence of a spatial dimension. They emphasised that, in their opinion, there would never be a general all encompassing database for GIS because the different families of applications demand distinct types of database support. This does not preclude the development of a tool for the management of integrity constraints within a range of applications as described in section 5.

As an intermediate theory to bridge the gap between conceptual spatial data models and physical implementations in existing SIS software, Hadzilacos et al (Hadzilacos and Tryfona 1994) defined a Geographic Relational Data Model that incorporated topological integrity constraints. They did not address the concept of incorporating layers or topological integrity constraints in relational theory, neither did they describe software to accomplish the automatic transition from conceptual to physical manifestations of the model although this was cited as a subject for research. This research would necessitate by its nature some form of data dictionary support. In (Marble 1990) the concept of an extended data dictionary was developed and a discussion is provided of the way its use can increase the integrity and long term useability of spatial databases.

Thus solutions put forward for integrity control fall into two categories; firstly control by means of object oriented methods for representing the structure and behaviour of objects used in geometric modelling, secondly spatial data management in relational DBMS extended to manage geometric objects. The notion of a data dictionary or repository is implicit, in the authors opinion, in the latter approach and should not be overlooked in the former as a means of defining the characteristics of spatial database objects and the methods which act upon them.

## 3. Why a repository?

In spatial information systems the traditional approach to database management has been to allow individual applications to supplement the set of capabilities offered by the underlying system architectures. In this way the operational needs of spatial data handling are satisfied, including integrity constraint checking. A more elegant way of handling integrity constraints is via a central repository. This removes the need for 'hard coding' integrity constraints into application code, thus reducing redundancy and making maintenance easier. Various researchers have suggested the repository of a means of removing the burden of managing GIS capabilities from the application (Marble 1990). Topics addressed include the management of integrity constraints (Cockcroft 1996b) and the management of GIS operations within the repository (Stefanakis and Sellis 1996). Chadwick (Chadwick 1995) has also proposed an architecture for managing spatial business rules, elements of which are incorporated here.

The ultimate aim of this work is improvement of data quality through the imposition of integrity rules on data entry. This will be done by incorporating constraints in data base schemas or enforcing user defined rules by means of triggering operations. This may necessitate a departure from existing techniques of data entry whereby spatial data is gathered first and topology added later. However, it is possible to envisage a situation where the rules are not imposed the time when spatial data is entered, but later, when topology and attribute data are added. At this stage rules could either be enforced or logged for later examination. In the approach described here the responsibility for data integrity lies within the scope of the DBMS rather than application programs. The rules from which triggers are derived are stored in the repository. A case study is described here which demonstrates an application that could use the repository to improve data quality as described above.

### 3.1. Constraints and Rules

The *Business rule* is a well established term in non spatial database literature. It refers to rules that are defined by the user or other major stakeholder in an enterprise that may be specific to a given application for example ' A pay rise cannot have a negative value'. Rules in the spatial information systems literature more frequently refer to expert system rules (Luo and Jones 1995)(Jones and Luo 1994). The rules discussed here would be more properly described as user rules and are in fact constraints. For a further discussion of spatial integrity constraints the reader is referred to an earlier paper (Cockcroft 1996a). One point of interest here is the idea that spatial integrity constraints may be defined in terms of attribute data. This issue was eluded to in the review paper by Gunther et al (Günther and Lamberts 1994) referenced earlier. An example of this was given in (Chadwick 1995) when referring to an SIS application for a pipe network:

*" A butterfly valve can only be connected to a pipe > 14 inches in diameter"*

Clearly although this rule will have to be implemented in both spatial and non spatial components of an SIS it is based on attribute data from the non-spatial component. The word *connect* has implication for the spatial component of the system since it implies a topological relationship. Topological relationships have been formally defined in (Egenhofer and Franzosa 1991). These would be used as a basis for describing the spatial relationships upon which the user rules would be defined.

### 3.2. Objectives

The preceding discussion outlined the shortcomings of current spatial information systems regarding their ability to enforce semantic integrity constraints. At the conceptual level the definition of spatial relationships and constraints thereon is well developed. However, at the implementation level this is not the case. Thus, two specific objectives of this work are to produce a GIS architecture which allows for:

- Topology to have some semantic information attached
- Spatial relationships to be defined on the basis of attribute data and be enforced in some way.

## 4. Approaches to integrating spatial and non-spatial data

Any discussion of spatial data integrity would not be complete without reference to emerging technologies. In particular those recent initiatives which have resulted in the ability to store spatial and non spatial data under the same architecture. The degree to, and the approach by, which this integration is achieved varies considerably. There seem to be two broad approaches; the management and storage of spatial data within extended relational databases and the use of object oriented techniques to integrate spatial and non-spatial data in a fashion that is seamless to the user. In the former category is Oracle7 Multidimension (Oracle 1995), Postgres (Stonebraker and Kemnitz 1991), and Illustra and Montage which are the commercial embodiments of Postgres. In the latter category are SIRO-DBMS (ONTOS)(Milne 1993), Smallworld (Yearsley et al. 1994), GODOT (Günther and Riekart 1993). The fact that the relational approaches incorporate object oriented techniques and the object oriented approaches are often owe much of their underlying structure to the relational model should not be overlooked. In fact a true object oriented spatial database would require the existence of an underlying formal foundation analogous to the relational model. As yet there is no formal object-oriented model. Efforts towards a consensus on what such a model should constitute are being made. It is generally accepted that such a model should support the constructs available in object oriented languages that is encapsulation, inheritance, class relationships and polymorphism

### 4.1. Existing solutions to managing spatial data integrity control

The use of an object oriented approach to system development means that objects, although referred to within application programs, need only be defined once. In addition the relationships in which they are involved and the constraints defined on them can also be defined. The prevailing technology is still the relational approach however and the solutions to user defined constraint management although well defined in the non-spatial area are not well developed in spatial information systems.

## 5. Repository system architecture

As mentioned earlier the purpose of the repository developed here will be concerned with the storage of design elements of spatial metadata with a view to automatically generating database schemas for the storage of spatial metadata. The design elements of interest are *dictionary* metadata - this describes the characteristics, relationships, and uses of data.

Part of semantic integrity control is the simple ability to state that a value must be include in a particular range as discussed in section 2. This is traditionally managed by domains. Domains are stored in the repository and specify constraints on valid values for attributes.

The pilot system described here is developed in a PC based client server environment. There is a market for this type of architecture in small to medium sized companies and government bodies as evidenced by recent literature (Frizzell and Cardno 1996)(Long and Barthelmeh 1996)(Henderson and Soon 1996)(Phare 1996). Additional support for this approach comes from the assumed environment in which data capture would occur. This work is designed to improve the accuracy of data entry and update in small to medium sized systems using desktop and digitiser data capture methods rather than large scale applications that would be more likely to use remote sensing or other automated data capture methods.

The system is designed to illustrate the usefulness of the repository. From the users point of view the integration is seamless. It has been suggested (Chadwick 1995) that a business rule processor is placed architecturally between applications and the business rule database so that software applications are isolated from the physical structure of the database, in this situation it would not be necessary for the applications to "know" how to interpret the business rules see Figure 1. This is the approach adopted here (see Figure 2)

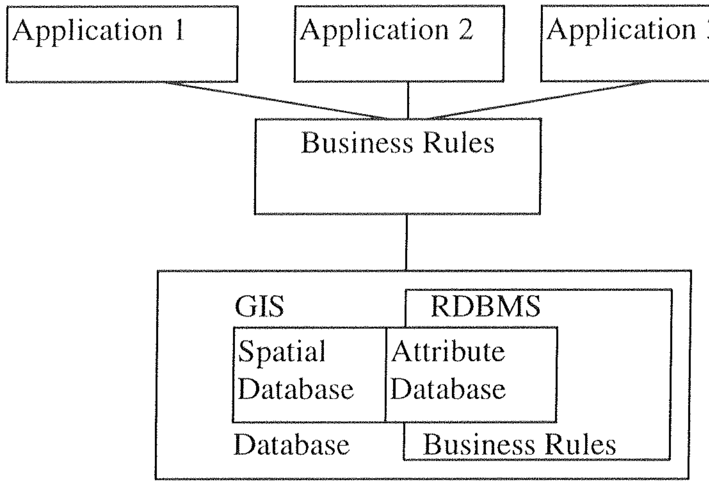


**Figure 1 Business Rules Application Architecture (Chadwick 1995)**

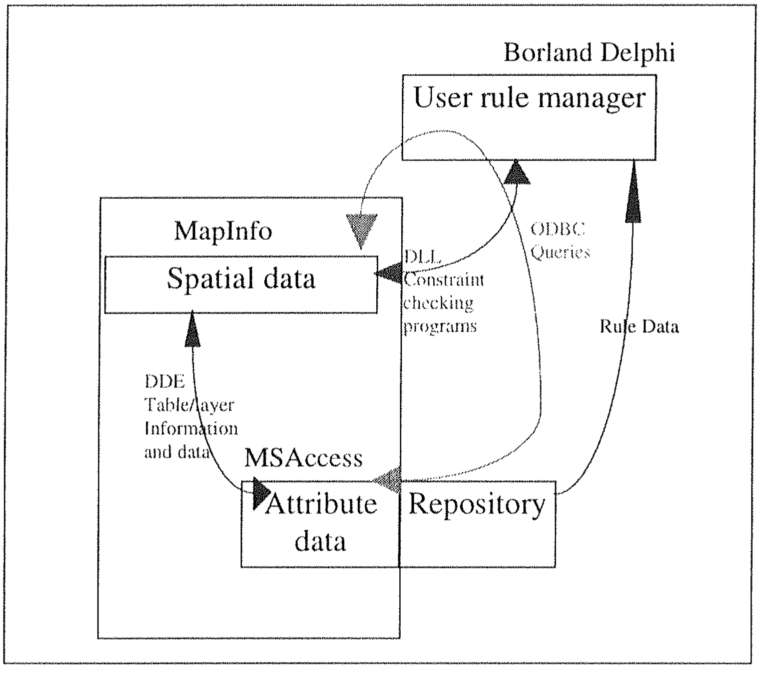


**Figure 2 Architecture for Spatial Information System with repository control**

By way of illustration a generic architecture for a GIS with repository control has been designed. MapInfo provides the interface with which the user will interact. The application data is stored in MapInfo and MSAccess. The reason that attribute data is not stored in MapInfo is that the facility provided by MapInfo doe not provide adequate Entity and Referential Integrity. Other GIS may be better in this regard. Application specific rules are stored in the repository but acted upon by the user rule manager. This configuration also allows for the sharing of rules between applications using DDE[1], ODBC[2] and DLL[3]. The reason that Delphi is used for developing the user rule manager is that it has a fast compiler and thus allows thousands of calculations to occur each time, for example, a region is added. In section 6 a case study is developed which shows how this architecture works.

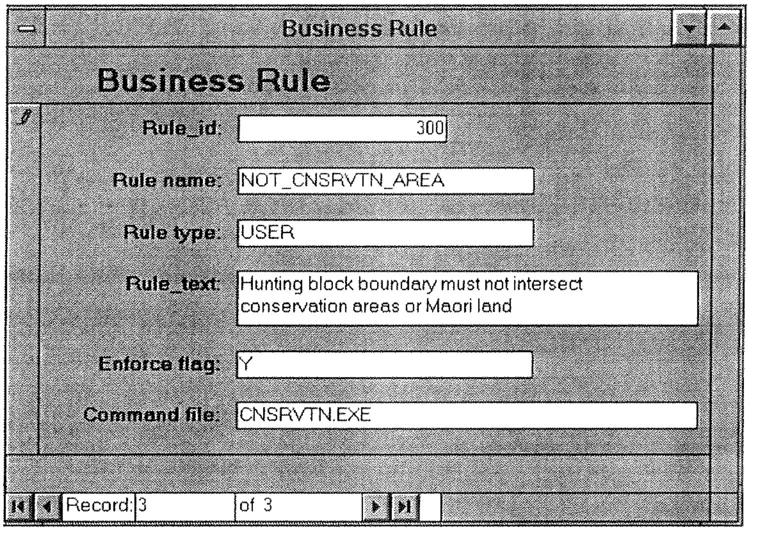


**Figure 3 Business rule data entry screen**

The business rule table in the repository is illustrated in Figure 3. Business rule names and their text description are recorded. The rule type and where it is enforced are also stored. Those not enforced by the application are enforced by the RDBMS. The file containing the executable file which will enforce the business rule is also stored. In the case study developed in section 6, the type of rules stored could include

- A hunting block must have an area greater than $10km^2$
- A hunting block must be bounded by recognisable features
- A hunting block must not encroach on residential, Maori or nature reserve land

This concept is further developed in the following section.

## 6. Case Study

To illustrate the operation of the repository and rule manager, A case study has been established. Stewart island is separated from New Zealand by Foveaux straight which has a minimum width of 27 Km. The Department of conservation (DoC) allows hunting

---

[1] DDE - Dynamic Data Exchange

[2] ODBC - Open database connectivity. A standard protocol for accessing information in SQL database servers

[3] DLL - Programs running in windows can share subroutines located in executable files called dynamic linked libraries

of Deer and Possum on the island but only according to certain regulations. At present DoC maintains hunting blocks which do not change over time (DOC 1995). This case study considers a hypothetical situation where a given area has become hunted out, or reserved for other reasons and it becomes necessary for DoC to alter the hunting block configuration.

Spatial Data would be stored in MapInfo and attribute data in MSAccess according to the architecture described in section 5. Rules concerning the hunting blocks would be stored in the repository which is also developed in MSAccess

When the user decides to modify the hunting blocks the first step is to enter the proposed boundary for the block spatially. An attribute table under the control of Access would pop up to receive attribute data. Upon receiving the information that this polyline represents the boundary of a hunting block, Access would check the rules relating to hunting block boundaries and send a request to Delphi to run a routine to ensure there were no other features at this location which would violate any of the boundary constraints. If such a violation occurred an error message would be issued as illustrated in Figure 4 and Figure 5. In this example the user has tried to modify the hunting block boundary to go through a conservation area.

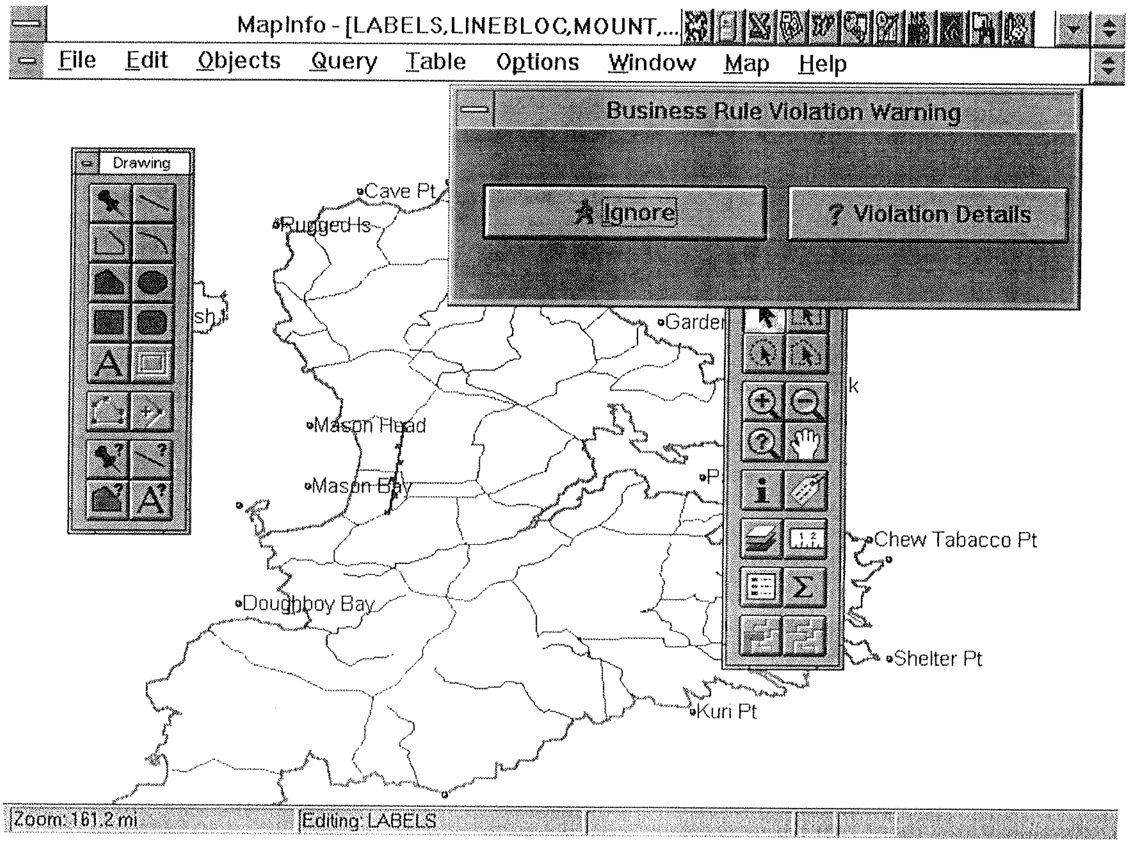


**Figure 4 Business rule violation example 1**

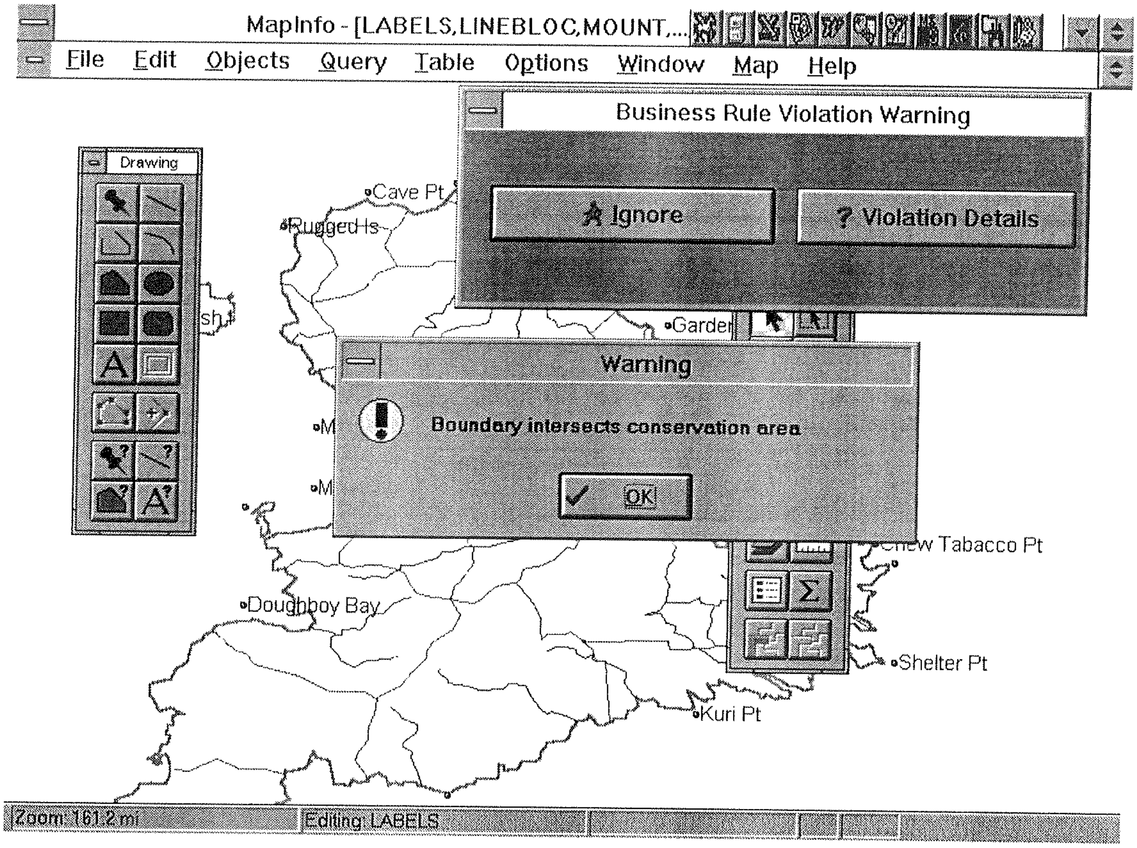


**Figure 5 Business rule violation example 2: Violation details**

## 7. Conclusion

This paper has described the development of a system architecture incorporating a spatial metadata repository analogous to the active data dictionary support offered in non-spatial DBMS. The challenge in developing such a system derives from the nature of spatial relationships, which are much more difficult to describe and manage in SIS than non spatial relationships in traditional DBMS. Early experiences with the prototype system architecture, which is not specific to the software described in the example, suggest that the implementation of integrity constraints based on user defined rules is feasible and thus represents a step forward in spatial data quality management.

## 8. References


1. Chadwick, A. "An Architecture to support business rules in custom GIS applications." *18th AM/FM International Conference*, 1047-1058.
2. Cockcroft, S. "Towards The Automatic Enforcement of Integrity Rules in Spatial Database Systems." *SIRC'96: Proceedings of the 8th Annual Colloquium of the Spatial Information Research Centre*, University of Otago, 33-42.
3. Cockcroft, S. 1996b. "The use of a metadata repository in spatial database development." *96/02*, Otago University, Dunedin.

4. DOC. 1995. "Stewart Island Hunting Information." *4*, Department of Conservation, Te papa atawhai, New Zealand.

5. Egenhofer, M. J., and Franzosa, R. D. 1991. "Point-set Topological Relations." *International journal of geographical Information Systems*, 5(2), 161-174.

6. Environmental Systems Research Institute, I. 1995. *Understanding GIS : the ARC/INFO method : self study workbook : version 7 for UNIX and OpenVMS*, GeoInformation International ; Wiley, 1995., Cambridge, U.K.

7. FGDC, F. G. D. C. 1994. "Content standards for digital Geospatial Metadata." , Federal Geographic Data Committee., Washinton, D.C.

8. Frizzell, P., and Cardno, A. "Integrating Successful Spatial Applications." *SIRC'96: Proceedings of the 8th Annual Colloquium of the Spatial Information Research Centre*, University of Otago, 212-218.

9. Günther, O., and Lamberts, J. 1994. "Object-oriented Techniques for the Management of Geographic and Environmental Data." *The Computer Journal*, 37(1), 16-25.

10. Günther, O., and Riekart, W.-F. 1993. "The design of GODOT: An Object-Oriented Geographic Information System." *IEEE Data engineering bulletin*, 16(3), 1-6.

11. Hadzilacos, T., and Tryfona, N. 1994. "Logical data modelling for geographic applications." *TR-94.01.1*, Computer Technology Institute, Patras, Greece, Patras, Greece.

12. Henderson, D., and Soon, S. "Extending GIS to practical applications in a utility business." *Conference on managing Geographic Information Systems for success*, Melbourne, Australia, 112-115.

13. Henderson, M. M. 1987. "The importance of data administration in information management." *Information Management Review*, **2**(4), 41-47.

14. Jones, C., and Luo, L. "Hierarchies and objects in a deductive spatial database." *Advances in GIS research: Proceedings of the 6th International Symposium on spatial data handling*, 589-603.

15. Long, M., and Barthelmeh, M. "Some Implications of GIS Software Choice." *SIRC'96: Proceedings of the 8th Annual Colloquium of the Spatial Information Research Centre*, University of Otago, 203-211.

16. Luo, L., and Jones, C.1995. "A deductive database model for GIS." Innovations in GIS 2, P. Fisherq, ed., Taylor and Francis, 33-42.

17. Marble, D. F. "The extended data dictionary: A critical element in building viable spatial databases." *11th annual ESRI user conference*.

18. Medeiros, C. B., and Pires, F. 1994. "Databases for GIS." *ACM SIGMOD record*, 23(1), 107-115.

19. Medyckyj-scott, D., Cuthbertson, M., and Newman, I. 1996. "Discovering environmental data: metadatabases, network information, resource tools and the GENIE system." *International Journal of Geographical Information Systems*, 10(1), 65-84.

20. Milne, P. 1993. "Geographical object oriented databases - a case study." *International Journal of Geographical Information Systems*, 7(1), 39-55.

21. Oracle. 1995. "Oracle7 MultiDimension™ Advances in Relational Database Technology for spatial Data Management." .

22. Phare, R. "Accessible GIS for Management and End Users." *Conference on managing Geographic Information Systems for success*, Melbourne, Australia, 76-81.

23. Stefanakis, E., and Sellis, T. "A DBMS repository for the Application Domain of Geographic Information Systems." *7th International Symposium on Spatial Data Handling (SDH '96)*, Delft, The Netherlands.

24. Stonebraker, M., and Kemnitz, G. 1991. "The Postgres next-generation database management system." *Communications of the ACM*, 34(10), 78-92.

25. Worboys, M. F. 1994. "Object-Oriented Approaches to Geo-Referenced Information." *International Journal of Geographical Information Systems*, 8(4), 385-399.

26. Yearsley, C., Worboys, M., Story, P., Jayawardena, P., and Bofakos, P.1994. "Computational support for spatial information systems: models and algorithms." Inovations in GIS - Selected papers from the First National Conference on GIS Research UK, M. Worboys, ed., Taylor & Francis, Keele, 75-88.